DOCUMENT RESUME

ED 129 907 TM 005 750

AUTHOR Froman, Robin D.
TITLE The Influence of Format Change on the Halo Effect of Student Ratings.
INSTITUTION Connecticut Univ., Storrs. Bureau of Educational Research and Service.
PUB DATE Apr 76
NOTE 18p.; Paper presented at the Annual Meeting of the National Council on Measurement in Education (San Francisco, California, April 19-23, 1976); For related document, see TM 005 744

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS Analysis of Variance; Effective Teaching; Factor Analysis; *Factor Structure; Higher Education; *Rating Scales; *Response Mode; Response Style (Tests); *Student Evaluation of Teacher Performance; *Teacher Rating; *Test Construction; Validity
IDENTIFIERS *Halo Effect

ABSTRACT

To determine whether placement of items on a teacher rating scale affects the factor structure underlying the scale and to determine whether changing the item format alters the ratings given, a 12-item, high-inference student rating scale was developed containing two global items pertaining to overall teacher effectiveness and 10 evaluative items pertaining to more specific teacher characteristics. Four different formats were constructed and randomly distributed to 363 college students within 27 university classes. The results of a principal components analysis suggest that the factor structure underlying the rating instrument is stable and not affected by changes in the placement of items. The nonsignificant results from a multivariate analysis of variance (MANOVA) provide additional support for the stability of the ratings. The MANOVA results in particular suggest that student ratings do not increase or decrease as a result of changes in item placement. The results of both analyses detract from the view that the halo effect may be influenced by the positioning of items. There is thus little reason to believe that preliminary items may set the stage for responses on following items. (RC)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

BUREAU OF EDUCATIONAL RESEARCH AND SERVICE

ED129907

U S DEPARTMENT OF HEALTH
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

ED 005 450

RESEARCH REPORT SERIES

SCHOOL OF EDUCATION
THE UNIVERSITY OF CONNECTICUT

The Influence of Format Change on the Halo Effect of Student Ratings

Robin D. Froman

University of Connecticut

A paper presented at the National Council on Measurement in Education, San Francisco, April, 1976.

The Influence of Format Change on the Halo Effect of Student Ratings

Robin D. Froman

University of Connecticut

Introduction and Review

One of the chronic problems of student ratings has been establishing content validity. Studies seeking to establish the technical adequacy of student ratings have generally supported the reliability of the measures, but validity remains controversial. It is not clear whether student ratings of teacher performance are actually reflections of teacher characteristics and teaching ability, or if they are in some part a reflection of student characteristics.

The halo effect has been claimed to cloud the interpretation of rating outcomes. The halo effect, in brief, is the tendency for raters to evaluate conceptually different characteristics as though they all represented the same thing. It is the tendency for prior information, or one impression or characteristic, to influence all other impressions. In other words, if an instructor receives a high rating on one characteristic, it is likely that he will receive a high rating on other characteristics.

Person-perception research has built a considerable argument for the existence of implicit personality theories, one possible cause or explanation for the halo effect. In a review of the research on the effects of implicit personality theory, Schneider (1973) suggests that

rating instruments are more affected by the preconceived personality theory of the rater than by the characteristics of the person being rated. Evidence is cited in the review to illustrate that similar trait structures emerge when different rating instruments are factor analyzed. These structures are virtually unaffected by the stimulus object, the familiarity of the rater with the ratee, differences in samples, or differences in situations (see, for example, Magoon and Price, 1972).

Within the literature pertaining specifically to teacher evaluation instruments, consideration has been given to the effect of implicit personality theory. Johnson, Rhodes, and Rumery (1973) suggest that teacher rating instruments do not necessarily furnish information about teacher performance, but rather, provide information about the acceptability of teacher characteristics and performance. Thus, teacher rating scales would appear to provide more information about student perceptions and implicit personality theories than about actual teacher behaviors. This notion is supported by the results of a study by Magoon and Price (1972). When comparing the results of principal component analyses of 21 sets of rating data between and within classrooms, the authors found that the major dimensions which emerged were very similar in structure. These researchers concluded that these striking similarities provided evidence that ratings reveal more about student preconceptions about the relationships between teacher traits than about the quality of instruction. In a later investigation of this possibility, Bausell and Magoon (1972) found a correlation of .67 between evaluations of teaching performance made on the first day of class and evaluations made at the end of the semester. In addition, the authors found that the changes in ratings

that did occur were predictable and related to changes in the grade a student expected to receive. If implicit personality theories are operating, one wonders whether stability estimates (of reliability) indicate consistency in teacher performance or consistency in students' beliefs.

Several studies have been conducted to determine what variables do or do not affect the bias in ratings which result from a rater's implicit personality theory. Schneider (1973) reported that occupation of the rater, caffeine arousal, and the perceived importance of the evaluation were found to influence the factor structures which emerged from analyses of rating instruments. Slater and Owen (1974) found that the major area of study of students affected the factor structure which emerged when students' ratings of hypothetical "ideal professor" were analyzed. The differences in the factors due to a student's major suggest that one's field of study may be a relevant predictor of the bias or implicit personality theory a student has.

Dixon and Dixon (1975) found this bias to be affected by the overall perception a student has of a teacher's competence. A comparison of teachers rated as "exceptionally good" versus teachers rated as "exceptionally poor" yielded two different sets of factor structures. The differences in the factor structures imply that students have different implicit personality theories with respect to good teachers and poor teachers.

A study by Follman, Lucoff, Small, and Power (1974) reports quantitative differences in ratings given to teachers as a result of format changes in scaling, coding, and response alternatives. There were, however, no analyses reported which would determine if there were changes

in the factor structures as a result of format changes. Thus the Follman et al. study shows evidence of changes in the mean ratings awarded based upon format changes, but does not determine if the dimensions underlying student bias change as a result of format changes.

The purpose of the present investigation was to determine if simple format changes would alter the existing halo effect and hence influence student ratings of teacher performance. The research was designed to find out whether the placement of questions on a rating scale is responsible for any halo effect. In particular, it was felt that global items pertaining to teacher effectiveness might provide a certain "tone" or "set" for students as they respond to subsequent questions. That is, having responded to initial items, the students might simply "follow suit" on the rest of the items on a scale.

The specific objectives of this study were twofold:

1. To determine whether placement of items on a teacher rating scale affects the factor structure underlying the scale; and
2. to determine whether changing the item format alters the ratings given.

Procedure

A twelve-item, high-inference student rating scale was developed containing two global items pertaining to overall teacher effectiveness and ten evaluative items pertaining to more specific teacher characteristics (see Appendix A).[1]

Four different formats were constructed, varying the items as follows:

1. The two global items were at the top of the rating instrument;

[1]For more information concerning the construction, validity, and reliability of the rating scale refer to, Student Ratings; What is the Frame of Reference? by A. Lolli and S. Owen, a paper presented at the Annual Convention of NCME, San Francisco, 1976.

2. The two global items were at the bottom of the rating instrument;

3. The two global items were absent from the rating instrument;

4. The two global items were the only items present on the form.

The four formats are shown in Appendices A through D. The four forms were randomly distributed to 363 students within 27 universitity classes. Undergraduate and graduate students from four academic areas[2] were represented in the sample.

Data from all classes were pooled and three principal components analyses were conducted to identify the factor structures underlying the ten evaluative items. Data from forms 1, 2, and 3 were considered in the factor analysis. A one-way MANOVA was applied to the data from Forms 1, 2, and 4, considering the two global items as the dependent variables. A second one-way MANOVA was performed on forms 1, 2, and 3, considering the ten evaluative items as the dependent measures (see Figure 1).

Results

Table 1 shows the results of the principal components analysis. The factor structures across the three forms containing the ten evaluative items were very similar. The strongest five items loading above .40 on Factor I were the same across the three forms, with slight changes in the order of item importance. The average percent of variance accounted for by Factor I was .39; the range from .34 to .42. Analyses of the data from form 3 yielded only two factors; the data from forms

---

[2]The academic areas represented in the sample were education, psychology, engineering and history.

1 and 2 yielded three factors. The two factor solution resulting from form 3 was apparently the consequence of collapsing the second and third factors which emerged on the other two forms into one factor. The collapsing of the factors was probably due to the eigenvalue criterion of 1.0. A third factor, quite similar to the third factor seen in forms 1 and 2, developed an eigenvalue of .95 and was not retained for rotation.

The slight changes in the relative importance of items in defining factors two and three on forms 1 and 2 can be explained as the result of a restriction of range on those items. In particular, items 5 and 7 had average variances of .37 and .21 respectively across the three forms. The average variances of the eight other evaluative items across the three forms compared in the factor analysis was .76.

- - - - - - - - - - - - - -

Insert Table 1 about here

- - - - - - - - - - - - - -

Two MANOVA's were conducted to determine if format changes would alter the ratings given. Neither MANOVA showed differences in the mean vectors of the dependent variables. The MANOVA analyzing the mean vectors of the ten evaluative items produced an F value of 0.83, $p < .68$, with degrees of freedom of 20 and 702. The analysis of the forms containing only the two global items showed an F value of .78, $p < .54$, with degrees of freedom of 4 and 710.

Conclusions

The results of the principal components analyses suggest that the factor structure underlying the rating instrument is stable and not affected by changes in the placement of the items. The nonsignificant

results from the MANOVA provide additional support for the stability of the ratings. The MANOVA results in particular suggest that student ratings do not increase or decrease as a result of changes in item placement. The results of both analyses detract from the view that the halo effect may be influenced by the positioning of items. There is thus little reason to believe that preliminary items may "set the stage" for responses on following items.

Although this research has not pinpointed the reasons for the halo effect, the negative findings may be viewed as a fortunate outcome. Researchers in instrument development may be somewhat more confident that general format revisions will not influence the technical adequacy of teacher rating scales.

APPENDIX A (Form 1)

The University of Connecticut Survey of Courses and Teaching

Department _____ Course # _____ Instructor _____

DIRECTIONS: Student ratings can be a powerful type of feedback to the instructor. Please be as objective as you can. DO NOT SIGN YOUR NAME.

Your Semester: ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8 ☐ Grad. Is this course in your major field? ☐ yes ☐ no

Expected grade in course: ☐ A ☐ B ☐ C ☐ D ☐ F ☐ PASS

Cumulative Grade Average:

☐ 1.0-1.5 ☐ 1.5-2.0 ☐ 2.0-2.5 ☐ 2.5-3.0 ☐ 3.0-3.5 ☐ 3.5-4.0

Percent of class attendance:

☐ 90-100% ☐ 75-89 ☐ 50-74 ☐ 25-49 ☐ 0-24

Time spent on this course compared to other courses of equal credit:

☐ considerably more ☐ somewhat more ☐ about the same ☐ somewhat less ☐ considerably less

How much have you learned from this course?

☐ almost nothing ☐ a little ☐ a moderate amount ☐ quite a bit ☐ a great deal

How would you rate this instructor in general, all-around teaching ability?

☐ poor ☐ less than adequate ☐ average ☐ good ☐ excellent

For the following ten questions, rate each item in two ways:

a. give your actual rating of the instructor; then
b. mark the second scale according to how you would prefer things to be. Of course, if you are satisfied, the second scale will be marked the same as the first.

From my observations, this instructor:	almost never	not very often	sometimes	most of the time	almost always
1. presents material in a clear & effective manner					
(preferred)					
2. stimulates interest					
(preferred)					
3. makes work assignments & student responsibilities clear					
(preferred)					
4. uses exam items which stress important aspects of the course					
(preferred)					
5. grades fairly and impartially					
(preferred)					
6. demonstrates overall organization					
(preferred)					
7. meets class regularly and on time					
(preferred)					
8. fulfills class objectives and obligations					
(preferred)					
9. is accessible to students both in and out of class					
(preferred)					
10. shows an interest in and concern for students					
(preferred)					

APPENDIX B (Form 2)

The University of Connecticut Survey of Courses and Teaching

Department _____ Course # _____ Instructor _____

Directions: Your instructors will not see these ratings until after grades are awarded. Please be as objective as you can, and DO NOT SIGN YOUR NAME.

Your semester: ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8 ☐ Grad. Is this course in your major field? ☐ yes ☐ no

Expected grade in course: ☐ A ☐ B ☐ C ☐ D ☐ F ☐ Pass

Cumulative grade average:

☐ 1.0-1.5 ☐ 1.5-2.0 ☐ 2.0-2.5 ☐ 2.5-3.0 ☐ 3.0-3.5 ☐ 3.5-4.0

Percent of class attendence:

☐ 90-100% ☐ 75-89 ☐ 50-74 ☐ 25-49 ☐ 0-24

Time spent on this course compared to other courses of equal credit:

☐ considerably more ☐ somewhat more ☐ about the same ☐ somewhat less ☐ considerably less

For the following ten questions, rate each item in two ways:

a. give your actual rating of the instructor; then

b. mark the second scale according to how you would prefer things to be. Of course, if you're satisfied, the 2nd scale will be marked the same as the 1st.

From my observations, this instructor...

	almost never	not very often	sometimes	most of the time	almost always
1. presents material in a clear & effective manner					
(preferred)					
2. stimulates interest					
(preferred)					
3. makes work assignments & student responsibilities clear					
(preferred)					
4. uses exam items which stress important aspects of the course					
(preferred)					
5. grades fairly and impartially					
(preferred)					
6. demonstrates overall organization					
(preferred)					
7. meets class regularly and on time					
(preferred)					
8. fulfills class objectives and obligations					
(preferred)					
9. is accessible to students both in and out of class					
(preferred)					
10. shows an interest in and concern for students					
(preferred)					

How much have you learned from this course?

☐ almost nothing ☐ a little ☐ a moderate amount ☐ quite a bit ☐ a great deal

How would you rate this instructor in general, overall teaching ability?

☐ poor ☐ less than adequate ☐ average ☐ good ☐ excellent

MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS 1963-A

APPENDIX C (Form 3)
The University of Connecticut Survey of Courses and Teaching

Department _____ Course # _____ Instructor _____

Directions: Your instructors will not see these ratings until after grades are awarded. Please be as objective as you can, and DO NOT SIGN YOUR NAME.

Your semester: ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8 ☐ Grad.

Is this course in your major field? ☐ yes ☐ no

Expected grade in course: ☐ A ☐ B ☐ C ☐ D ☐ F ☐ Pass

Cumulative grade average:
☐ 1.0-1.5 ☐ 1.5-2.0 ☐ 2.0-2.5 ☐ 2.5-3.0 ☐ 3.0-3.5 ☐ 3.5-4.0

Percent of class attendance:
☐ 90-100% ☐ 75-89 ☐ 50-74 ☐ 25-49 ☐ 0-24

Time spent on this course compared to other courses of equal credit:
☐ considerably more ☐ somewhat more ☐ about the same ☐ somewhat less ☐ considerably less

For the following ten questions, rate each item in two ways:
a. give your actual rating of the instructor; then
b. mark the second scale according to how you would prefer things to be. Of course, if you're satisfied, the second scale will be marked the same as the first.

From my observations, this instructor...	almost never	not very often	sometimes	most of the time	almost always
1. presents material in a clear & effective manner					
(preferred)					
2. stimulates interest					
(preferred)					
3. makes work assignments & student responsibilities clear					
(preferred)					
4. uses exam items which stress important aspects of the course					
(preferred)					
5. grades fairly and impartially					
(preferred)					
6. demonstrates overall organization					
(preferred)					
7. meets class regularly and on time					
(preferred)					
8. fulfills class objectives and obligations					
(preferred)					
9. is accessible to students both in and out of class					
(preferred)					
10. shows an interest in and concern for students					
(preferred)					

APPENDIX D (Form 4)

The University of Connecticut Survey of Courses and Teaching

Department _____ Course # _____ Instructor _____

Directions: Your instructor will not see these ratings until after grades are awarded. Please be as objective as you can, and DO NOT SIGN YOUR NAME.

Your semester: ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8 ☐ Grad.

Is this course in your major field? ☐ yes ☐ no

Expected grade in course: ☐ A ☐ B ☐ C ☐ D ☐ F ☐ Pass

Cumulative grade average:

☐ 1.0-1.5 ☐ 2.5-3.0
☐ 1.5-2.0 ☐ 3.0-3.5
☐ 2.0-2.5 ☐ 3.5-4.0

Percent of class attendance:

☐ 90-100%
☐ 75-89
☐ 50-74
☐ 25-49
☐ 0-24

Time spent on this course compared to other courses of equal credit:

☐ considerably more
☐ somewhat more
☐ about the same
☐ somewhat less
☐ considerably less

How much have you learned from this course?

☐ almost nothing ☐ a little ☐ a moderate amount ☐ quite a bit ☐ a great deal

How would you rate this instructor in general, overall teaching ability?

☐ poor ☐ less than adequate ☐ average ☐ good ☐ excellent

BIBLIOGRAPHY

Bausell, R. B. and Magoon, J. The persistence of first impressions in course and instructor evaluations. A paper presented at the annual meeting ot the American Educational Research Association, Chicago, April, 1972.

Dixon, J. K. and Dixon, J. Comparison of factor structures of student ratings of exceptionally good teachers versus exceptionally poor teachers. A paper presented at the annual meeting of the Northeastern Educational Research Association, Ellenville, N. Y., October, 1975.

Follman, J., Lucoff, M., Small, L., and Power, F. Kinds of keys of student ratings of faculty teaching effectiveness. A paper presented at the annual meeting of the American Educational Research Association, Chicago, April, 1974.

Johnson, H. C., Rhodes, D. M., and Rumery, R. E. The assessment of teaching: the end of the drunkard's search? A paper presented at the annual meeting of the American Educational Research Association, New Orleans, February, 1973.

Magoon, A. J., and Price, J. R. Rating dimensions of course and instructor characteristics: the eye of the beholder. A paper presented at the annual meeting of the American Educational Research Association, Chicago, April, 1972.

Schneider, D. J. Implicit personality theory: a review. Psychological Bulletin, 1973, Vol. 79 (5), 294-309.

Slater, J. K., and Owen, S. V. Departmental differences in student perception of the "ideal" professor. A paper presented at the annual meeting of the Northeastern Educational Research Association, Ellenville, N. Y., October, 1974.

FIGURE 1

Description of MANOVA's to Determine if Format Changes Alter Ratings Awarded

Form 1 Two global items on <u>top</u>	
Form 2 Two global items on <u>bottom</u>	TWO GLOBAL ITEMS
Form 4 Two global items <u>only</u>, evaluative items absent	

Form 1 Ten evaluative items on <u>bottom</u>	
Form 2 Ten evaluative items on <u>top</u>	TEN EVALUATIVE ITEMS
Form 3 Ten evaluative items <u>only</u>, global absent	

TABLE 1

Summary of Three Principal Components Analyses of Ten Items Contained on Forms 1, 2, and 3 of the Teacher Rating Scale

Form 1		Form 2		Form 3	
Item	Loading	Item	Loading	Item	Loading
		Factor I			
2.	.808	1.	.892	1.	.911
1.	.801	2.	.848	2.	.866
8.	.786	8.	.842	6.	.816
6.	.785	6.	.836	8.	.647
10.	.638	10.	.412	10.	.507
3.	.468	3.	.405	5.	.471
		Factor II			
5.	.744	5.	.694	3.	.631
4.	.714	7.	.633	7.	.615
3.	.400	4.	.557	4.	.569
		3.	.541	9.	.543
				10.	.532
				5.	.530
		Factor III			
9.	.810	9.	.931		
7.	.704	10.	.527		
		5.	.403		

TABLE 2

Means and Standard Deviations of the Four Formats of the Teacher Rating Scale

	Form 1 N=120		Form 2 N=118		Form 3 N=125		Form 4 N=121	
	$\bar{X}$	sd	$\bar{X}$	sd	$\bar{X}$	sd	$\bar{X}$	sd
Item								
1.	4.27	.80	4.37	.88	4.26	.84		
2.	4.22	.94	4.16	.95	4.11	.94		
3.	4.41	.93	4.39	.86	4.45	.87		
4.	4.38	1.10	4.02	1.23	4.15	.97		
5.	4.74	.49	4.60	.63	4.65	.69		
6.	4.26	.91	4.19	1.05	4.16	.92		
7.	4.78	.55	4.82	.43	4.86	.37		
8.	4.47	.66	4.47	.82	4.43	.80		
9.	4.62	.66	4.58	.74	4.54	.65		
10.	4.63	.59	4.58	.77	4.57	.75		
Global 1	4.04	.96	3.91	1.00			3.99	.84
Global 2	4.24	.76	4.27	.87			4.29	.83